The Future of Data-Driven Decision Making: Exploring the Governance Models of Data Collaboratives and Their Relative Success

HONORS RESEARCH THESIS

Presented in Partial Fulfillment of the Requirements for graduation "with Honors Research Distinction in Public Affairs" in the undergraduate colleges of The Ohio State University

By

Josh Badzik

Shaun Loftin

The Ohio State University

April 2021

Project Advisor: David Landsbergen, J.D., Ph.D.

Abstract

There is a growing interest in using data-driven decision making in public policy. One response to this need is data collaboratives, which seek to fill the gap between data aggregation and public utilization of said data. Data collaboratives are a platform upon which different kinds of public and private data are collected, stored, and managed among private and public stakeholders to share data and conduct analysis. Different types of emerging data collaboratives include private intermediaries for data collection and public partnerships with smart city programs. Studying these collaboratives can provide insights into the future on how the government uses data by exploring its interaction with citizens in creating and implementing policy. In this paper, we will review various data collaboratives and look at their organizational leadership, governance approaches, mission statements, and successes/shortcomings. The research team used a mixed-methods approach by first conducting interviews to develop a more robust understanding of the nature of the problems and possible solutions. These conclusions were then validated through a survey and follow-up interviews. The results of the survey showed that many data collaboratives experience similar challenges - such as bureaucratic limitations and funding shortages - as they attempt to produce deliverables. Many data collaboratives are often narrowly focused on a policy issue such as transportation, healthcare, or infrastructure; therefore, looking at other examples of collaboratives in one city could contrast with the governance approach of another. Learning more about the successes and barriers of existing data collaboratives can help interested cities and regional partners build a comparable model.

Acknowledgements

We would foremost like to acknowledge our research advisor Dr. David Landsbergen for answering a cold email from two undergraduate students looking for a challenge. Taking two years of time out of his busy schedule is an incredible ask and we are thankful for your support.

We would also like to acknowledge Angie Westover-Muñoz for her remarkable work on technical analysis alongside her teaching schedule, as well as for her assistance leading initial interviews.

Finally, we would like to thank the Mid-Ohio Regional Planning Commission (MORPC) and all of the outstanding groups that participated in our survey and interviews throughout this study. This research wouldn't have been possible without the energetic support of the data collaborative, smart city, and academia communities. The willingness to advance knowledge on successes and shortcomings of data collaboratives has made this an exciting project to work on since day one.

Table of Contents

The Future of Data-Driven Decision Making: Exploring the Governance Models of Data Collaboratives and Their Relative Success

1 Introduction

The terms “Smart Cities” and “Open Data” have become 21st-century buzzwords among local governments wanting to advance their infrastructure using data. There is a growing interest in using data-driven decision-making to create effective public policy (Walravens & Ballon, 2013). These terms help to describe efforts that local governments have been initiating to collect, analyze, and disseminate data relevant to their constituents.

1.1 What is a data collaborative?

One response to this need for a better way to work with data and citizens is data collaboratives. Data collaboratives are a new kind of organization in which different kinds of public and/or private data are collected, stored, and managed among private and public stakeholders to share data and conduct analysis (Walravens & Ballon, 2013). These collaboratives are extremely useful tools for public officials; they can be used to communicate vast amounts of demographic, economic, and other statistical data to constituents (Solomon, 1997). Furthermore, data collaboratives are tools that, at the hands of public officials, can be used to inform policy and governing decisions (Susha et al., 2017).

Data collaboratives are an improvement upon governments’ repeated attempts to work with people and data. Currently, governments’ major efforts are working on providing “open data” (Sieber & Johnson, 2015). “Open Data” is about making public, and increasingly private, data available in technical formats so that it can be easily communicated and analyzed (Meijer & Potjer, 2018). Typically, this data is found on websites, such as the federal government’s Data.gov (GSA, 2020) open data website. Stakeholders can simply go to a website, search for the data they want, and download it 24/7, without making any requests for the data. Open Data is an improvement on even earlier innovations like FOIA or Open Records requests where citizens must go through a cumbersome and costly request process that often resulted in a denial of that request (Landsbergen, 2004). Open Data services involve proactively providing data to the public without waiting for a FOIA or request.

The problem is that while Open Data has been around for more than a decade, we are still "stuck" in the mindset that just providing this public data on a website will be enough to ensure that it will be useful and used. It turns out that there are still many barriers to using data that are not solved by merely providing public data on a website (Zuiderwijk, 2015). The next generation of making data available will move from this old passive approach to taking a positive, engaged service-oriented approach, one of which is the "data collaborative." Studying these collaboratives may provide insights into the future of how the government uses data and interacts with the public in making policy (Sicilia et al., 2016). Local governments are increasingly turning toward collaboratives to guide their decision-making in relation to policy program planning, as well as in times of crisis and public emergency.

Now that governments have had some time experimenting with data collaboratives, it is time to collect information and to begin to learn from experience how they operate and what works. The proposed research is the first attempt to gather empirical data on the 'governance' of data collaboratives; namely, how data collaboratives are structured, managed, and funded. Results of this work could lay the foundation for further in-depth research on these subtopics (Susha et al., 2017) as well as some general recommendations to data collaboratives.

1.2 Research Questions

An important question in the establishment of any data collaborative is the "who"; the "who" owns the data, the "who" in dividing responsibilities, and "who" drives the agenda. Our first research question explores how data collaboratives are governed and how the public-private partnership is defined. These partnerships can be explicitly defined using charters or informally through a series of personal connections. While there are many different styles of governance, their approach and implementation of "who" assumes various shared responsibilities directly affects their success.

Another important question in the establishment of any data collaborative is the "how"; the "how" this partnership provides value, the "how" success is defined, and "how" success is measured. Our second research question explores which governance approaches are successful. Each partnership uses different measures of success, whether it is a qualitative measure of data

usefulness or quantitative number of deliverables, and how success is defined is key to evaluating their impact. Some governance models and approaches have proven to be more capable to define and reach these measures of success than others.

While this work is descriptive and exploratory and does not involve any hypothesis-testing, based upon prior work in graduate capstone courses, the research team expects a number of findings: 1) One of the primary barriers to the efficacy of data collaboratives is the need for better management of the information within the local government (Landsbergen, 2021). 2) Short-term incremental projects will have the greatest success rate. 3) "Government-controlled" data collaboratives have the highest likelihood of success. 4) Active engagement of citizens provides the greatest gains in legitimacy and trust in the results. 5) There will be relatively more analysis (a limited set of researchers drawing conclusions) as compared to doing "sense-making" where various sets of stakeholders work together to find shared meaning in the conclusions (Dervin, 1998).

The overall goal is to learn about how to govern and manage data collaboratives so that: 1) data is better utilized in making decision-making; 2) the decision-making improves because multiple stakeholders are not only consulted on what they value but are now actively involved in analyzing and making sense of the data; and 3) citizen understanding of, and trust in, government increase.

1.3 Outline of Writing

This paper outlines how data collaboratives are governed and what kinds of governance approaches are successful. We will first provide a working definition of the term "data collaborative." Though there are several ways in which one can define a data collaborative, ours will be applicable in the context of local government usage. Then, we will provide an overview of data collaborative governance; we will discuss "who" owns the data and "who" drives the agenda, as well as the "how" their strategic goals are accomplished through a number of volunteers. We will then provide our research methodologies used to explore local governments' use of data collaboratives. Next, we describe successful approaches to governance, highlighting both qualitative results from a survey and quantitative results confirming our survey findings

through follow-up interviews. Subsequently, we will explore opportunities for city and regional governments to adopt data collaboratives. We will evaluate the feasibility of these opportunities, and then define the barriers to entry that prevent many cities and regional governments from reaching full adoption. Finally, we will provide recommendations for local governments that are seeking to create and successfully implement data collaboratives.

2 What is a data collaborative?

The term “data collaborative” can have different connotations dependent on one’s sector, region, and exposure to these groups. We established a broad, working definition of a data collaborative that we used throughout our interviews and survey, ensuring we were referring to the same type of organization. Our definition and understanding of these groups was refined by New York University’s The Governance Lab’s resources regarding data collaboratives (Simone-Noveck, 2020). We define the term generally to err on the side of learning about many organizations rather than inadvertently excluding organizations that might be interesting. Relying on their division of data collaboratives into six different unique types, we then began to hypothesize which activities, challenges, and funding problems data collaboratives may face.

2.1 Working Definition of a Data Collaborative

We define data collaboratives to be “platforms through which data is collected, stored, and managed among private and public stakeholders for the purposes of data sharing and sense-making”, we acknowledge that there are several working definitions within the industry as well as various sub-classifications within each main group (Dervin, 1998). While our working definition is expansive and inclusive of a lot of groups, data collaboratives are adapting, learning from one another, and electing to take on new roles and activities. Furthermore, the activities that collaboratives undertake are necessarily expansive and inclusive as they adapt to each collaborative’s operational growth over time.

2.2 Smart Cities

“Smart city is an innovative form of a city that utilizes information and communication technology to help the quality of life of its citizens” (Zulkarnain et al., 2019). Smart cities are a form of data collaborative used to connect big data with IT governance. Under this definition, a collaborative is a governing body that utilizes information and communication technology to help the quality of life of its citizens. “Integrated technology can help provide information to its citizens, improve service efficiency and also the welfare of its people” (Zulkarnain et al., 2019). Though technology is commonly developed to advance healthcare, transportation, energy, and water needs, such advancements are not necessarily limited to this area. Alternatively, some big

data experts characterize data collaboratives in the context of social governance. Under this definition, data collaboratives are a process for modernization that seeks to redefine the development of human society. Through various transformations of societal governance, collaboratives provide measures for adapting and rethinking the way government interacts with citizens as they work, study, and live.

Our interest and perspective of data collaboratives was initially shaped by an interest in data-driven policy making within Columbus and their smart city initiatives, specifically our work with the Mid-Ohio Regional Planning Commission (Murdock, 2021). MORPC has a working relationship with The Ohio State University, the John Glenn College of Public Affairs, Smart Columbus, local governments in the Mid-Ohio area, and other planning commissions nationally. The commission runs a number of committees including, but not limited to: Transportation Policy Committee, Sustainability Advisory Committee, and the Regional Data Advisory Committee.

We sat on a subcommittee within the Regional Data Advisory Committee called the Data Policy Needs Survey & Toolkit Working Group. This group aimed to benchmark data sharing, security, and collaboration efforts within Mid-Ohio governments to understand shortcomings and create effective "toolkits". Toolkits in the context of data refer to instructional guides or deliverables given to groups to help bring them up to a benchmark in terms of data security, privacy, etc. The concept of benchmarking current data efforts within Mid-Ohio inspired a willingness to pursue this curiosity, effectively beginning our research inquiry, but on a more national scale.

2.3 Benefits of Data Collaboratives

Data collaboratives are utilized as a tool for removing and reducing barriers between governmental organizations and their attempts to engage with citizen participants. Collaboratives serve as information hubs that provide increased transparency and improved decision making, therefore decreasing tension that can arise from a lack of such public involvement. Though certain forms of data released by collaboratives can create unnecessary nuance and dubiety, when crafted and delivered with careful precision, these platforms can result in substantial and

effective improvements to the relationship between government-affiliated and non-government-affiliated data experts.

A data collaborative can also be a partnership created for the explicit purpose of exchanging and making available data for use by each of the member parties for its own public policy-making and problem-solving goals. These partnerships can involve any number of actors, often ranging from government agencies to private companies to research institutions. The partners that form the data collaborative do not have to overlap in their sphere of influence, allowing for enhanced data availability among all parties. In the context of local governments, data collaboratives are a technical tool that policymakers can utilize to improve public services for their constituents.

2.4 Types of Data Collaboratives

There are a variety of different collaborative types as defined by New York University's GovLab including Application Programming Interface (API), Data Pooling, Intelligence Product, Prizes & Challenges, Research Partnerships, and Trusted Intermediaries (Verhulst, 2020). As defined in Table 1, we chose these definitions because we thought they were inclusive of all different types of activities and vary in the accessibility of data they provide to the public, deliverables, and upkeep needed.

Table 1: New York University's GovLab Data Collaborative Definitions

Activity	**Definition**
Data Pooling	Organizations agree to create a unified presentation of datasets as a collection accessible by multiple parties.
Prizes & Challenges	Organizations make data available to participants who compete to develop apps; answer problem statements; test hypotheses and premises; or pioneer innovative uses of data for the public interest and to provide business value.
Research Partnerships	Organizations engage directly with public-sector partners and share certain proprietary data assets to generate new knowledge with public value.
Intelligence Products	Organizations internally develop data-driven analyses, tools, and other resources, and release those insights to the broader public.
Application Programming Interface (API)	Organizations provide open access to certain data assets, enabling independent uses of the data by external parties.
Trusted Intermediary	Third-party actors support collaboration between private-sector data providers and data users from the public sector, civil society, or academia.

Our research primarily focuses on data collaborative examples most in line with Research Partnerships, Trusted Intermediary, or Data Pooling. Data collaboratives that fall under the category Research Partnerships, for example, often include a major local university that helps drive the analysis. Data collaboratives that fall under the category of Data Pooling utilize a combination of public and/or private data across multiple partners for collaborative use. While

the type of data collaborative can depend on the established goals of the group, the activities performed with the data directly depend on the type of data collaborative.

2.5 Activities of Data Collaboratives

After forming partnerships and collecting the data, another aspect that uniquely identifies data collaboratives is the activities they conduct. “Activities” is a loose term that describes the day-to-day functions of the group including data analysis, their deliverables, networking opportunities, and more. “Sense-making activities” is a term that refers to activities relating to analysis and dissemination of data with an organization (Dervin, 1998). Some data collaboratives with larger funding structures choose to perform data analysis and provide various deliverables (such as data dashboards, policy recommendations, or publications) in-house, whereas others choose to outsource it to the broader community. Other data collaboratives focus their efforts to be more networking-focused and use their energy to connect professionals to public issues.

One’s activities are integral to the identity of a data collaborative. For example, some data collaboratives may focus on engaging with constituents or community partners to help with project ideation, data collection, and data analysis. Other data collaboratives may alternatively choose to serve governments or private clients more directly by completing data tasks in-house without constituent interaction. Charters of data collaboratives help define the mission of the group, which then helps determine the activities that best execute said mission given their resources.

2.6 Challenges of Data Collaboratives

There are many different kinds of challenges that data collaboratives face not only at their inception but also years into their operations. Our list of challenges has been inspired by a combination of our initial interviews with collaboratives and literature review. One major challenge that data collaboratives face is that they are still being formed and identified. There is no universal definition of a “data collaborative”, only working definitions based on the models of successful cities thus far. Another major challenge is securing a stable funding model that will ensure the longevity of the collaborative. Data collaboratives aren’t an overwhelmingly popular

topic among citizens, so the efforts to create these have to originate with public leaders. Without vocal citizenry support, it is difficult to justify funding for a team of salaried in-house software developers and GIS specialists.

2.7 Funding of Data Collaboratives

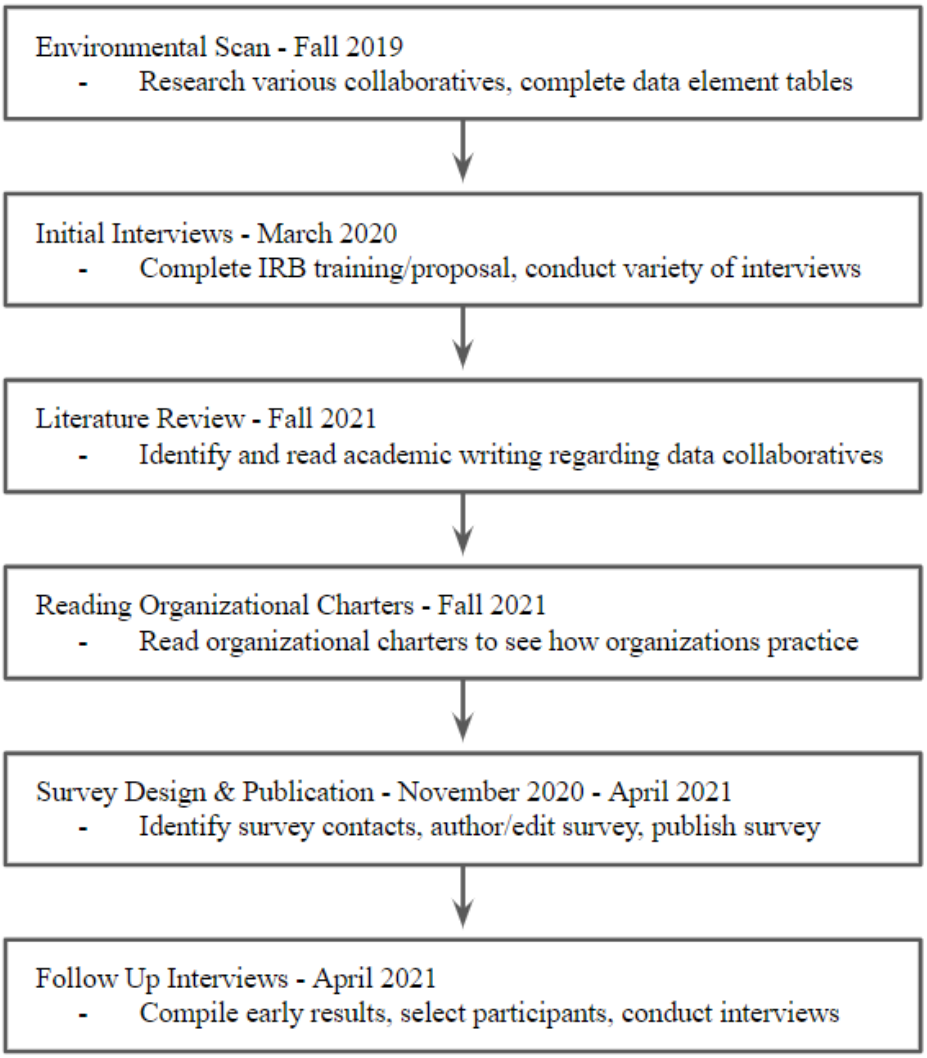
Funding for data collaboratives can come from a wide variety of sources. Though funding can come from city budget revenues, this is much less common. Most commonly, the collaboratives are funded through state and federal grants that are allocated to local governments. Additional sources of funding include partnerships with private companies, non-profit organizations, as well as both private and publicly-funded colleges and universities. City budgets are an indicator of the priorities of a community's governing officials. Notably, data collaboratives are funded through a mixture of sources other than local revenue streams and often partner with neighboring cities or regions to increase their yield .

Cities are incentivized to invest time and resources into data collaboratives because of the long-term benefits they will yield from participation. While required initial inputs include data, manpower, and financial resources, long-term outputs include expansive shared regional data and the potential to create data toolkits for use in public services. Shared regional data can allow for adjacent areas to strategize on the policy as well as improve both the quality and efficiency of public services. Data toolkits are guides that help municipalities and departments work towards centralized data standards in security, metadata, and more-help improve data utilization and assist local government in improving data deliverable quality across the board. While data toolkits are a convenient resource for data collaboratives, they are most effective in use with larger city/regional data collaboratives.

3 Methodologies

With a foundational understanding of what data collaboratives have to offer, we needed to develop research methodologies that continued to explore the “who” and “how”. We needed to understand “who” these groups are nationwide, “who” is driving the agenda, and “who” these groups looked up to in their founding. We also needed to understand the “how” these data collaboratives are run by understanding their activities, governance models, and how they provide value to stakeholders. In order to answer these questions, the research employed a mixed-method approach. Interviews were conducted with a number of different kinds of data collaboratives to gain a better understanding of the different kinds of activities data collaborative might engage in, the opportunities and challenges they face, and the various governance approaches taken to manage these opportunities and challenges. After a qualitative analysis of these interviews resulted in working conclusions were developed, and in combination with a literature review, survey questions were developed to validate the lessons learned. The flowchart below, Figure 1, shows our research design and timeline.

Figure 1: Research Design Flow Chart



3.1 Environmental Scan of Data Collaboratives

The research began with an environmental scan of data collaboratives around the country. The scan used the work from a prior experiential capstone class developing a local data collaborative. The environmental scan included a review of literature on data collaborations outside the U.S. to identify unique perspectives but the focus was on domestic data collaboratives. Building from these two sources, a semi-structured questionnaire was used to interview the data collaboratives. The interviews ended with a question asking that data

collaborative if there were any additional data collaboratives they know of and that might be interesting for the research team given our stated needs.

Our scan consisted of a strategic review of the data collaboratives reported on the NYU Data Collaborative website. Information gathered from respective websites and were put into a tabular format designed by our research team. These data element tables, as shown in Table 2, helped summarize key information of each collaborative we could later use to compare and contrast.

Table 2: Environmental Scan Sample Data Element Table

Sample Data Element Table
● Name / Dates ● Contact Info / URL ● NYU website categorizations ● Charter ● How are they organized and run? ● Who are stakeholders? Audience? ● Kind of participants ● Collect data from >1 organizations ● Observed Data Types ● Traditional analysis of data ● Community understanding of what data means ● Funding Sources ● Do they have measures of success? ● What are your pain points? ● What do you want to do next? ● Have others failed? Who? ● Whom do you follow? ● Summary comment. What is the big picture here and where do they fit in?

We first examined the purpose and structure of each collaborative. Then, we investigated the activities and specific “sense-making” actions undertaken by each institution. Throughout this examination, we made careful note of key stakeholders in both of these processes, both

internal and external. Additionally, we sought to gauge the measures of success with which each collaborative could measure and/or report their progress to said stakeholders. We also outlined the origins and legacies of each collaborative, referring to who they looked to as a model and what other collaboratives they have since provided inspiration and/or guidance. These stories helped us to better identify the signs of failure that indicate the low likelihood of success for a collaborative in the early stages of development.

3.2 Initial Interviews

We connected with and gathered data from collaborative organizations using a two-phase approach. The first phase was to conduct interviews to develop a more robust understanding of the nature of the problems and possible solutions. Prior to beginning our first phase of interviewing data collaboratives, we began the OSU IRB application process. We completed relevant IRB training modules for Human Subjects Protection and Responsible Conduct of Research. Next, we drafted an IRB proposal in compliance with federal rules which was approved in Spring 2020.

Using the questions listed in Appendix A, we conducted detailed phone interviews with data collaboratives in Denver (DRCOG), Raleigh-Durham (TRDC), Pittsburgh (WPRDC), City of Tulsa, Los Angeles (LACity) in order to gather background information regarding the purpose and structure of their respective regional data collaborative. Additionally, we wanted to gain a better understanding of the different kinds of activities data collaborative might engage in, the opportunities and challenges they face, and the various governance approaches taken to manage these opportunities and challenges. After a qualitative analysis of these interviews, working conclusions were developed and in combination with a literature review, survey questions were developed to validate the lessons learned.

3.3 Literature Review

Our literature review consisted of a detailed search and analysis of journal articles relating to the topic of data utilized in the context of public policy. Using the Ohio State

University Libraries virtual catalog, we specifically looked for publications that featured key words such as data collaboratives, smart cities, and data governance as shown in Table 3.

Table 3: Search Terms Used to Find Relevant Academic Literature

Key Terms
Data collaboratives, Smart cities, Data governance, Data governance models, Regional data, Data sharing, Open data, Data impact, Data partnership, Big data, Sense-making, Data policy, Geographic information systems, Public value, Public private platforms, Data bureaucracy

After compiling an extensive database of written works, we reviewed the articles that best-related to the scope of our research and took note of the findings reported. In our review, we discovered the surprising lack of published work pertaining to data collaboratives and solidified our understanding of their operation as being in the early stages of development.

3.4 Reading Organizational Charters

Another data source useful in our research was obtaining documents that chartered the organization and described how they governed. We also obtained data sharing agreements as they would make clear the responsibilities of the various actors. Many data collaboratives operate on a set of charters, either formal or informal, that help identify the goals, division of responsibilities, and identity of the data collaborative. The charters better define the initial structure of each organization, as well as point to key stakeholders in the entire process of the collaborative's actions.

3.5 Survey

During the second phase of connecting with and gathering data from collaborative organizations, we sought to validate the findings from our interviews and what we had read in the review of the literature. We conducted an IRB-approved research survey using an Ohio State managed version of Qualtrics and followed a rigorous brainstorming, writing, and revising process that took a couple of months. We established primary goals for the survey and based our

revisions to meet these goals: less than 10 minute estimated response time and minimize long-form questions. With the understanding that response time is a major barrier for getting and completing responses, we removed and consolidated questions to prioritize our time. Reducing the amount of long-form and opting for more choice-based questions aimed to improve rate while streamlining analysis. Our final survey asked 22 questions and had an estimated completion time of 9 minutes (Appendix B).

Research participants were identified from a number of sources. Our research team gathered contact information by examining websites of target data collaboratives. We sought to identify particular persons we could send the survey to in order to make sure that the correct person is filling it out and to ask someone to take on ownership to increase the chances of it being completed. In particular, we requested that the survey be completed by the person who "best understands the governance of the data collaborative." Furthermore, we looked at the networks of those data collaboratives we had already interviewed. We also identified additional data collaboratives that were registered with the NYU Data Collaborative website. Finally, Google searches were used to make sure that the six different categories of data collaboratives were represented to make sure that there were a sufficient number of data collaboratives for each of the categories developed in our research. The following search parameters were used, but not limited to those listed in Table 4.

Table 4: Search Terms Used to Find Data Collaboratives

Key Terms
Community Data, Data Center, Data Collaborative, Data Partnership, Hackathon, Health Collaborative, Innovate, Metro Area Data Collaborative, Neighborhood Indicators, Open Data, Regional Data, Regional Data Collaborative, Urban Data, Urban Institute, Urban Research
Types
Public, Private, Nonprofit, Public-Private Partnership, Data Intermediary, Data Pooling, Research Partnership, Intelligence Products, Application Programming Interface (API), Trusted Intermediary
Locations
Northwest, Southwest, Northeast, Midwest, California, Colorado, Connecticut, District of Columbia, Georgia, Hawaii, Idaho, Illinois, Kansas, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Nebraska, Nevada, New Mexico, New York, North Carolina, North Dakota, Ohio, Oregon, Pennsylvania, Tennessee, Texas, Utah

These terms were deliberately chosen as they represent a wide spectrum of organizations that identify as data collaboratives. The term "data collaborative" isn't always used in the title even if applicable; thus, using terms such as "regional data" and "community data" often revealed data collaboratives that branded themselves in a more niche manner. The search terms for types remained true and consistent with the New York University definitions used. The various states and regions under our location search terms aimed to represent different areas of the United States. Concatenating the name of a larger state or metropolitan area and a key term in a Google Search often yielded a couple of results in which we found individual contacts and added them to our database.

We then sent personalized emails to these contacts identifying the purpose of the research and the research team. We identified the person we were seeking to contact (the person who best

understands the governance of the data collaborative" and if there was someone better to answer those questions. We asked that they provide that contact information to us. The second email was the actual invitation to participate, which made clear that the survey was anonymous and what participant rights were granted. As shown in Appendix C, a series of three follow-up emails asking potential participants to complete the survey were also distributed. Potential participants had the opportunity to unsubscribe to survey completion reminder emails at any time.

We provided an opportunity for participants to learn of the results of the survey by having them provide their contact information. We made clear that this was optional, and also asked if participants would be willing to be contacted for follow-up interviews. Early results and conclusions of the survey were compiled into a document to be shared with selected participants who would be completing a post-survey follow up interview. These early results shared with interviewees and comprehensive results will be discussed below as well as sections 4.2 and 4.3.

3.6 Follow-Up Interviews

An overwhelming majority of respondents indicated their interest in follow-up interviews. Interviews were conducted with consenting groups to follow up on our early results (Appendix D). These participants were asked to corroborate or contradict our results based on their own lived experiences and findings. Data collaboratives of diversified types, activities, and experiences were selected to get a breadth of perspectives. For those selected, their availability and the availability of the research team were coordinated to find a common time. Along with the early results, a document describing our research procedures and the interviewee's rights were sent in advance in an informed consent guide for their review.

We conducted 30-minute interviews in which we asked different respondents their thoughts on our early survey results in accordance with their work in this field. These interviews provided more direct conversations that focused on not only the work in their data collaborative but how their activities, funding structures, challenges, etc., align with or differ from the seeming majority of our results. For members of the research team who couldn't be in attendance, the interviews and brief summarizing notes were recorded for them to view later.

4 Results

Utilizing different quantitative and qualitative methods of gathering data, we've been able to find consensus that there are many barriers to public participation, common challenges to secure and sustain funding, and a willingness to expand to better serve their constituents with more activities. Our environmental scan gave us a strong initial understanding of the field of data collaboratives and the diversity in organizational identities that fit that term "data collaborative". We then applied our initial hypotheses in a survey that helped confirm or deny the commonalities we expected. While our survey response rate was too low to allow us to make statistically significant conclusions about the population, our follow-up conversations with respondents did provide some validation of our findings and help us understand the nuances in the survey results.

4.1 Environmental Scan

Given our personal familiarities with collaborative data efforts in the Columbus area, we were expecting a majority of data collaboratives in our environmental scan to be affiliated with either local government or academica in some way. Our environmental scan using datacollaboratives.org challenged our original assumption as we found many more private and issue focused collaboratives than we were expecting. Reading publicly available charters allowed us to explore what governance models were out there and how they were organized/run. Information regarding their specific activities, challenges, and funding wasn't often available publicly, necessitating the need for a survey.

4.2 Survey

We sent out emails to 195 contacts asking to complete the survey and our Qualtrics survey received 31 responses total. After removing partial and duplicate responses, we received 23 usable responses yielding a completion rate of 74%. The survey was open for about four weeks and contacts who hadn't yet opened or completed the survey were sent up to two reminder emails. Raw data from Qualtrics was cleaned using Microsoft Excel and a codebook was created to convert text-based answers into numerical choice values (Appendix E). Our analytical approach was driven by our two research questions: the "who" of the data collaborative and the "how" their data collaboratives operate. After looking at our answer frequencies and early

results, we explored how responses differed based on sector, activities, and more. These curiosities were tested using two sample wilcoxon tests but we didn't have enough respondents for a statistically significant p-value. While we know that the response rate was low (need to calculate and report this somewhere) we still report on the responses that were received.

4.2.1 Activities

Given that activities are paramount to their identity and how they connect with clients & constituents, respondents were asked to identify which main activity their data collaborative best identifies with using New York University's six main GovLab Data Collaborative Definitions. Using the definitions reported in Table 1, Table 5 shows that Intelligence Products and Trusted Intermediary were the most common main activity among respondents.

Table 5: Main Activity Survey Frequencies

Main Activity	**Frequency**	**Percentage**
Data Pooling	4	17.39%
Prizes & Challenges	0	0.00%
Research Partnerships	2	8.70%
Intelligence Products	7	30.43%
API	0	0.00%
Trusted Intermediary	7	30.43%
Other	2	0.00%

The lack of identification with Prizes & Challenges and API as their main activity didn't come as a surprise, as these are normally secondary activities within a well-established group. While some respondents with larger full-time staff numbers appeared to take on all of the listed activities, that wasn't always the case; thus, suggesting that having a larger staff doesn't always imply that they take on more activities (but rather continue to specialize). Respondents were then asked to select all various activities that apply to their data collaborative, ranging from data collection, to analysis, to collaboration.

Table 6: Kinds of Activities Survey Frequencies

Kinds of activities	**Frequency**	**Percentage**
Collect data	17	73.9%
Analyze data	21	91.3%
Disseminate data	20	87.0%
Disseminate analysis	21	91.3%
Work with Clients on Projects	18	78.3%
Curate data	18	78.3%
Work with Citizens on Projects	12	52.2%
Clean data	17	73.9%
Create ways to combine data	19	82.6%
Other	5	21.7%

The results from Table 6 were a bit surprising in terms of the amount of activities data collaboratives are able to accomplish on average. 91.3% of respondents indicated that they participate in data analysis and dissemination, a clear commonality among data collaboratives of varying sector and type. Through our interviews we know that funding limits activities, so it comes as a surprise that more collaboratives aren't currently working with citizens as an inexpensive means to work on projects. Conversely, private data collaboratives may not serve a specific constituency and may face data-privacy barriers (ex. HIPAA) that prohibits from easily collaborating with citizens. This specific table raised more questions about the role of volunteers, staff, or consultants in activities that would be later answered in follow-up interviews.

4.2.2 Challenges

According to our survey respondents, funding – or a lack thereof – is the greatest challenge facing existing data collaboratives. In fact, more than 82% of collaboratives surveyed reported that funding was a significant challenge facing their organization. This metric is consistent with our data surrounding funding sources, which suggested that collaboratives utilize a vast array of funding sources to support the needs of their organizations. Our conversations with collaboratives also helped to shed light on the constraints that a lack of funding places onto the operation of a collaborative; limited staffing is a significant constraint that results from this

situation. A limited staff not only hurt's the collaborative's ability to complete tasks and meet deadlines by overwhelming employees with high workloads, but it also confines future growth by reducing available resources for recruitment and hiring.

Table 7: Types of Challenges Survey Frequencies

Challenges	**Minor**	**Significant**	**Total**	**Percent**
Funding	3	16	19	82.6%
Sufficient technical resources	13	3	16	69.6%
Bureaucratic limitations	7	9	16	69.6%
Guidance/Knowing how	9	1	10	43.5%
Time/Deadlines	9	8	17	73.9%
Lack of stakeholder support	8	2	10	43.5%
Finding volunteers	8	1	9	39.1%
Lack of data collaborator support	9	6	13	56.5%
Other	0	5	5	21.7%

Table 7 emphasizes that in addition to funding, collaboratives reported many other barriers to success. Bureaucratic limitations, deadlines, and a lack of collaborator support were all noted as challenges by more than half of respondents. Each of these challenges' present additional difficulties that data collaboratives must overcome to be successful in making their work useful to external organizations. Furthermore, a lack of technical resources presents internal barriers to success within a collaborative's ability to complete projects and was reported to be at least a minor problem for nearly 70% of respondents.

4.2.3 Funding

Our survey analysis reported that Foundation support was by far the greatest funding source for the collaboratives surveyed, with 30% of all funding coming from this source. This figure stands in sharp contrast to our preconceptions of collaborative funding, which we predicted to be overwhelmingly funded by government grants. Grants, however, accounted for only 25% of total funding sources, with federal grants representing nearly 70% of that figure. Though a range of other fee structures – such as fee for service, crowdfunding, and subscription

fees – were represented within the survey sample, they each accounted for only small percentages of the overall sources of funding. This response, as shown in Tables 8 and 9, regarding funding sources suggests that collaboratives are experimenting with a diverse mix of revenue streams and have not yet been successful in unifying around a common method of securing funding.

Table 8: Types of Funding Survey Frequencies

Funding	**Average**	**Std. Deviation**	**Min**	**Max**
Local government revenue	7%	12.92	0	50
State grant	4%	8.84	0	30
Federal grant	17%	28.95	0	100
Crowd funded	2%	7.13	0	35
Public Private partnership	3%	12.24	0	60
Privately funded	3%	7.14	0	30
University	7%	19.15	0	90
Subscription or membership fees	9%	22.50	0	70
Foundation support	30%	31.29	0	100
Other	8%	17.44	0	70
Fee for service	11%	25.55	0	100

Table 9: Number of Sources of Funding Survey Frequencies

Number of Sources	**Frequency**	**Percentage**
1	4	17.4%
2	9	39.1%
3	3	13.0%
4	4	17.4%
5	1	4.3%
6	0	0.0%
7	1	4.3%

4.2.4 Bureaucracy vs. Public Support

The relationship between data collaboratives and public beneficiaries requires collaboratives to carefully navigate the bureaucratic process while simultaneously seeking to acquire and maintain public support. The bureaucracy component of this struggle comes from the burdensome policy and political constraints that weigh on governmental organizations. Government agencies inherently face more public scrutiny due to their public nature, and therefore have many more measures of accountability than do private or nonprofit organizations. Public support for data collaboratives arises from the benefits that they provide citizens and government constituencies. Though data collection and analysis pose costs to consumers, the overall net benefit of these services within a community lead to a level of support from community members. Therefore, collaboratives must balance the responsibility of adhering to and navigating the bureaucratic process with their underlying mission of creating value for the public – who will, in turn, support their existence.

4.3 Interviews

We conducted six initial interviews in March 2020 and six follow-up interviews in April 2021. These two different sets of interviews, a year apart, had different purposes and learning outcomes. Using the questions in Appendix A, our initial set of interviews provided us information regarding specific data collaboratives, their governance models, and allowed us to build a professional network within this industry. We concluded interviews by asking what data collaboratives they looked towards for inspiration and we asked them to connect us via email. These individuals were later contacted for an interview or to complete the survey.

Using the guide found in Appendix D, six follow-up interviews were conducted in April 2020 in which respondents were asked to help confirm or challenge some of the early results of the survey based on their lived experiences. These interviews lasted anywhere between 30 to 60 minutes and were largely conversational.

In response to results 1-2 in Appendix D, interviewees weren't surprised by these results. Some noted that their ability to have volunteers is near impossible due to strict data privacy

standards; alternatively, others noted that their ability to utilize volunteers has grown recently leveraging the interests of students looking to gain exposure in this field.

In response to results 3-4 in Appendix D, taking a look into the frequencies of funding sources seemed surprising to many. In about half of our interviews, the response to the amount of federal grant funding and foundation support is "where can I apply for those?" Talking about funding, specifically which grants and foundations, among data collaboratives seems to be a taboo subject due to their competitive nature. Many noted a lack of resources on where to find these grant options and easily apply for them. One interviewee noted that grant applications "are focused on innovation and expansion rather than paying staff" and that regular operational costs are paid through more stable, recurring funding sources.

Some data collaboratives are fortunate to have governmental or public support that helps build value that is recognized through local government revenue. Other collaboratives rely on charging fees from either clients or users for services as their recurring revenue. One interviewee noted that "people want to pay for a project, not the staff for the project". While both funding models have been proved to be successful, one's funding model is often based around the needs of their stakeholders.

In response to results 5-7 in Appendix D, interviewees overwhelmly agreed even if they didn't find them applicable to their collaborative. The result of nonprofit being the most common is consistent with Trusted Intermediary being one of the most popular main activity. Interviewees noted that successful governance models need to have a "disinterested third party" for maintaining objectivity and non-profits are often viewed as trustworthy holders of data.

In response to results 8-9 in Appendix D, interviewees overwhelmly agreed that funding and time and deadlines were all major barriers. Bureaucratic limitations can vary the most out of the three major barriers listed depending on how they are run. For example, one interviewee expressed frustration not being able to sign contracts or documents needed on behalf of their partnering university.

In response to 10 in Appendix D, interviewees commented on bullet points that they specifically identified with. Some noted that completing more analysis and building models for better data quality control are priorities within the next year. Most noted that building better relationships with data collaboratives is something they can improve on and a goal they hope to attain in the near future.

4.4 Limitations in Results

While these results represent a variety of different sectors, types, and ages of data collaboratives, it is important to note there are some limitations in our results. Our results cannot be extrapolated for all data collaboratives as our survey may have a considerable response bias. For example, private surveys who are particularly protective of their organizational processes may not have felt comfortable taking the survey. Another important consideration for our survey is that beyond frequencies of responses, many of our attempted correlations didn't have a high enough response rate to be statistically significant. For example, we hypothesized that there may be a correlation between sectors and challenges (in which public data collaboratives face different challenges than private) but our p-value wasn't convincing beyond just a coincidence. Our results are a step in identifying a breadth of governance models used and challenges faced but should be applied with caution.

5 Conclusion

Our research goal was to learn more about the successes and shortcomings of existing data collaboratives to help build a model that can be applied to interested cities and regional partners. Learning about the mistakes of past attempts allows organizations to better serve their stakeholders. However, as our research evolved, we realized that there is no "one-size-fits-all" solution for both building and sustaining the work of a data collaborative. The challenges of building versus sustaining are unique as varying problems arise in pursuing initial funding versus continuing to justify funding, pursuing more activities, and incorporating citizen engagement. The results of our survey and our interviews are more 'suggestive' than 'determinative' of things that data collaboratives might want to consider. After analyzing the results of the survey in full, as well as considering the suggestions of all survey respondents who participated in the post-survey interview process, we have formulated several key recommendations that provide insight into the steps that prospective and current data collaboratives should take to guarantee a stable existence.

5.1 Recommendations for Building a Data Collaborative

For anyone looking to establish a data collaborative, regardless of whether you are public or private or a non-profit, authoring charters is a strong first step. Charters allow a forming group to set initial expectations and structure, set quantifiable goals that can be used to justify funding, and provide security should the team change. Another strong first step is to explore grant funding, in particular, in addition to other recurring sources. We've found that grant funding is primarily focused on development and expansion rather than recurring operational costs. Utilizing these opportunities when starting out can be key for quick organizational growth. Lastly, the next step is to accurately understand one's community and possible community partners. Benchmarking constituents' needs through surveys, roundtables, and forums can be a great way to gauge community wishes while sparking public interest in your work. Identifying community partners, experts, and advisory partners within your community can also help build collaborative relationships. Especially starting out, data collaboratives may have a limited staff and may need to rely on these partnerships to outsource some activities.

5.2 Recommendations for Sustaining a Data Collaborative

After a data collaborative has been established, several key steps should be taken to ensure that the collaborative continues to function effectively. First, collaboratives should prioritize both tangible short-term and long-term target benchmarks. This ensures that the organization is meeting the week-to-week needs of its clients while simultaneously considering the overall financial stability of the organization. Our open-ended questions in the survey asking what they would recommend frequently underlined the importance of collaboratives should maintain consistent and detailed communication between themselves and their stakeholders. This allows the organization to continually recognize and continue to meet the needs of all community members who are involved in the functions of the data collaborative. In addition to stakeholders, collaboratives should communicate with one another in order to maintain a network that can provide support across various regions of the U.S. Our post-survey interviews with collaboratives highlighted the importance of both formal and informal conversations that provide additional perspectives about both baseline operations and long-term strategic planning. Finally, collaboratives should set and strive to accomplish long-term organization goals. This will safeguard the growth potential of the data collaborative and align the work of staff members to the enduring success of the aforementioned goals. Because data collaboration is a dynamic process, collaboratives should continue to keep an eye on how to continually develop, whether it is by understanding new relevant technologies, being open to other ways that data collaboratives, or taking advantage of the natural learning process as they grow and mature.

5.3 Limitations in Recommendations

While these recommendations are relevant for the current state of data collaboratives, these groups are rapidly evolving and barriers for entry or sustainability will evolve simultaneously. For example, as these collaboratives become more proven to provide value to local governments, funding or bureaucratic support may become less prevalent issues. Another consideration, as mentioned in our results section, is our sample size and demographic. Our sample size isn't large enough to be truly representative of all data collaboratives and our recommendations should be applied with such consideration. Due to their competitive interest

and privacy concerns, our response doesn't accurately represent the needs of private groups as it is skewed towards public and nonprofit collaboratives.

5.4 Next Steps In Data Collaborative Research

We believe that future research in this area could take a deeper look into how data collaboratives interact with each other. As the field of data collaboratives grows and evolves with new additions, the ability to work with other data collaboratives to share successes is of the utmost importance. Informal and formal networks between collaboratives are a key support system ensuring success for many organizations; a deeper analysis of the structure and resources provided by these networks could be extremely valuable in supporting both the establishment and growth of future collaboratives. Organizations like the National Neighborhood Indicators Partnership have been able to formulate such a model in which various data collaboratives can symbiotically work together (NNIP, 2021). Another direction that future research into data collaboratives could take is looking into how citizen engagement and involvement has evolved over time. Data collaboratives will continue to grow in size and to new regions. The ability for the average citizen to access public data and records will only grow more over time. As data collaboratives grow, they begin to gain more public awareness and capacity to engage with volunteers.

In conclusion, data collaboratives offer a new and promising approach for local governments to create data-driven policy while better connecting with their constituents. Data collaboratives are expanding and continuously being explored by interested cities as an opportunity to develop more informed, data-driven decisions. These collaboratives provide a unique opportunity to assist policy makers in bringing public services into the 21st century using existing technologies. However, there are specific barriers to consider before forming a data collaborative. Funding, bureaucratic limitations, and fast deadlines are some obstacles we've found that these organizations often run into. The benefits of using a data collaborative as a means to better understand one's constituency outweigh these possible challenges. This community is willing to learn and eager to share their successes and shortcomings with one another. The field of data collaboratives will continue to thrive with the ongoing enthusiasm to help one another and should continue to be explored.

Bibliography

Dervin, B. (1998). Sense‐making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, *2*(2), 36–46.

GSA. (2020) “U.S. Government Open Data.” *Data.gov*, www.data.gov/.

Landsbergen, David. (2004). Screen level bureaucracy: Databases as public records. *Government Information Quarterly*, *21*(1), 24–50.

Landsbergen, David. (2021) Public Affairs 7960: Information Management Capstone, Spring Semester.

Meijer, A., & Potjer, S. (2018). Citizen-generated open data: An explorative analysis of 25 cases. *Government Information Quarterly*, *35*(4), 613–621.

Murdock, William. (2021) “Mid-Ohio Regional Planning Commission.” *MORPC*, www.morpc.org/.

NNIP. (2021) “National Neighborhood Indicators Partnership.” *Neighborhood Indicators*, www.neighborhoodindicators.org/.

Sicilia, M., Guarini, E., Sancino, A., Andreani, M., & Ruffini, R. (2016). Public services management and co-production in multi-level governance settings. *International Review of Administrative Sciences*, *82*(1), 8–27.

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, *32*(3), 308–315.

Simone-Noveck, Beth. (2020). *The Governance Lab*. The GovLab. https://thegovlab.org/.

Solomon, P. (1997). *Discovering information behavior in sense making. III. The person*. 12.

Susha, I., Janssen, M., & Verhulst, S. (2017). *Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development*. Hawaii International Conference on System Sciences.

Verhulst, S. (2020). *Data Collaboratives*. The GovLab. https://datacollaboratives.org/.

Walravens, N., & Ballon, P. (2013). Platform Business Models for Smart Cities: From Control and Value to Governance and Public Value. *IEEE Communications Magazine*, 8.

Zuiderwijk, A. (2015). Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government Information Quarterly*, 12.

Zulkarnain, N., Kosala, R., Ranti, B., & Supangkat, S. H. (2019). Big Data Governance for Building A Smart Cities. *2019 International Conference on ICT for Smart Society (ICISS)*, 1–5.

Appendix

Appendix A: Initial Interview Questions

Governance of Data Collaboratives

Interview Protocol
<March 2, 2019>

Previous to Interview

- Identify your interviewee and title for the target data collaborative. (A list of target collaboratives will be developed before the interview process begins.)
- Having made contact, use the following sentence in searching for that person. "We are looking for the best person in the organization who can explain how the data collaborative is governed and managed."
- Identify charter and/or governing documents for the organization on the website. If it cannot be found, ask for them at the end of the interview.
- Call target interviewee and use recruitment script.
- Follow-up by sending the recruitment script, informed consent form and the list of interview questions in an email to the interviewee.

Interview Questions

- [Introduce yourself but before you begin ask the interviewee a question, ask whether the interviewee agrees to have the interview recorded. If so, turn on the recording on the rev.com app. Then, ask the interviewee if they have read the informed consent form. Ask them whether they agree to participate in the research (you should be able to get their assent on the video recording.] Just in case, the video recording does not take place, not the time, date of their assent.

- Please provide in a couple of sentences how you would summarize the purpose and goals of the data collaborative.

- How was the data collaborative formed? How has it evolved over time? In answering your question could you discuss:
    - What are the main activities of the collaborative? How is the work carried out?
        - Probe on what activity they carried out in both: 1) data collection; and 2) analysis <----> sensemaking continuum. 3) Listen for any other kinds of activities.

        - "Sense-making" involves the "collaborative effort of different kinds of stakeholders to understand what the data means."
    - How is the data collaborative governed and managed?
    - Who are the major stakeholders in your work?
        - Probe for whether public, private, nonprofits or universities are involved.
    - What kinds of people and groups participate in your work?
        - Probe for the categories of people they use in describing which people and groups participate.
        - You can also listen for whether there are responsibilities or functions that might typically be taken on by different kinds of people or different kinds of groups within the collaborative.
    - What are your sources of funding for the data collaborative?
    - How do you measure success in your work?
    - Whom do you collect data from?

- Barriers and Drivers of Success
    - What challenges do you face in your work?
    - What do you need to be more successful?

- Are there any other data collaboratives, data pools, or general collaborations that inspired the data collaborative?

- <If could not identify charter or governing documents, request them during or at the end of the interview.> Identify charter and/or governing documents for the organization. If it cannot be found, ask for them at the end of the interview.

Appendix B: Survey Questions

Data Collaborative Environmental Scan

Start of Block: Introduction

Q1
The goal of this nation-wide survey is to share information about data collaboratives, specifically their governance and management.

'Data collaboratives', most generally, are organizations that collect data from multiple sources."

Your contributions to this survey are anonymous and will not be shared with others except in a summary manner so that you cannot be identified.

Before we begin, we must obtain your informed consent to participate.

The survey will take about nine minutes to complete.

Your responses will be automatically saved in case you get interrupted. You can continue later using the same link.

Thank you for participating.

If you would like to receive the results of this survey, there is an opportunity at the end of the survey to provide your name. We may also find it useful to follow-up on our results with some interviews. If you would be willing to provide your name, please let us know.

Q2
INFORMED CONSENT
This is a consent form for research participation.
It contains important information about this study and what to expect if you decide to participate.

Your participation is voluntary.
Please consider the information carefully. Feel free to ask questions before making your decision whether or not to participate. If you decide to participate, you will be asked to select the "I agree to participate" box to proceed to the survey.

PURPOSE
Researchers at The Ohio State University are conducting this survey to study data collaboratives.

RESEARCH PROCEDURES
You will be asked a series of questions about your work with data collaboratives.

DURATION
The survey will take approximately 9 minutes.

RISKS
There are no foreseeable risks for participating in this research beyond those associated with everyday life.

BENEFITS
There are no benefits to you as a participant other than to further research in public management.

CONFIDENTIALITY
Your data will be protected with a code to reduce the risk that other people can view the responses. We will work to make sure that no one sees your survey responses without approval. But, because we are using the internet, there is a chance that someone could access your online responses without permission. In some cases, this information could be used to identify you.

PARTICIPANTS RIGHTS Your participation is voluntary, and you may withdraw from the study at any time and for any reason. You may skip questions you feel uncomfortable answering. If you decide not to participate or if you withdraw from the study there is no penalty or loss of benefits to which you are otherwise entitled. There are no costs to you or any other party. This research has been reviewed according to The Ohio State University procedures governing your participation in this research. The study was found to be acceptable according to applicable state and federal regulations and University policies designed to protect the rights and welfare of participants in research.

CONTACT
Professor David Landsbergen at The Ohio State University are conducting this research. David Landsbergen can be reached at (614) 795-6002 for questions, concerns or complaints about the study, or if you feel you have been harmed as result of study participation. You may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at the Ohio State University at (800) 678-6251 if you have questions or comments regarding your rights as a participant in the research.

CONSENT

I have read this form and agree to participate in this study. If you agree to participate, please select "I agree to participate" and press the submit button.

- ○ I agree to participate (4)
- ○ I do not agree to participate (5)

Skip To: End of Survey If INFORMED CONSENT This is a consent form for research participation. It contains important infor... = I do not agree to participate

End of Block: Introduction

Start of Block: Overview Questions

Q3 Approximately, how many full-time equivalent employees do you have?

_____

Q4 Approximately, how many full-time equivalent volunteers do you have?

_____

Q5 In what year was your data collaborative formed? (YYYY)

_____

Q6 What best describes the main activities of the collaborative? Please pick the best answer.

- ○ Data pooling -- Organizations agree to create a unified presentation of datasets as a collection accessible by multiple parties. (4)

○ Prizes & Challenges -- Organizations make data available to participants who compete to develop apps; answer problem statements; test hypotheses and premises; or pioneer innovative uses of data for the public interest and to provide business value. (5)

○ Research partnerships -- Organizations engage directly with public-sector partners and share certain proprietary data assets to generate new knowledge with public value. (6)

○ Intelligence products -- Organizations internally develop data-driven analyses, tools, and other resources, and release those insights to the broader public. (7)

○ Application Programming Interfaces (API) -- Organizations provide open access to certain data assets, enabling independent uses of the data by external parties. (8)

○ Trusted intermediary -- Third-party actors support collaboration between private-sector data providers and data users from the public sector, civil society, or academia. (9)

○ Other (11) _____

Q7 What kinds of activities do you engage in? (Please check all that apply.)

☐ Collect data (1)

☐ Analyze data (2)

☐ Disseminate data (3)

☐ Disseminate analysis (4)

☐ Work with Clients on Projects (5)

- [ ] Curate data (8)
- [ ] Work with Citizens on Projects (6)
- [ ] Clean data (9)
- [ ] Create ways to combine data (10)
- [ ] Other (7) _____

End of Block: Overview Questions

Start of Block: Organizational Governance

Q8 Which sector best describes your group?

- ○ Public (1)
- ○ Private (2)
- ○ Public Private Partnership (PPP) (4)
- ○ Non-profit (3)
- ○ Other (5) _____

Q9 What are your sources of funding for the data collaborative? Please enter the ***approximate*** percentage to the right of each revenue source. For example, if crowd-funding is 10% of your funding, enter 10 in the appropriate box. They should total to 100.
Local government revenue : _____ (4)
State grant : _____ (5)
Federal grant : _____ (6)
Crowd funded : _____ (7)
Public/private partnership - joint assumption of providing resources and assuming risk : _____ (8)
Privately funded - receive dollars or resources from the private sector : _____ (9)
College/University : _____ (11)
Subscriptions or Membership Fees : _____ (12)
Foundation Support : _____ (13)
Other : _____ (10)
Total : _____

Q10 Who are the stakeholders in your work? (Please check all that apply.)

- ☐ Public Entity/Governmental Body (5)
- ☐ Citizens (6)
- ☐ Private Entity/Company (7)
- ☐ Non-Profits (8)
- ☐ University (9)
- ☐ Research Group (10)
- ☐ Public (13)

☐ Customers (12)

☐ Other (11) _____

Q11 Whom do you collect data from? Please enter the ***approximate*** percentage to the right of each data source. For example, if requests or contributions specifically from citizens are 10% of the data you use, enter 10 in the appropriate box. The total contributions should total to 100.
Public Entity/Governmental Body : _____ (4)
Citizens : _____ (5)
Private Entity/Company : _____ (6)
Non-Profits : _____ (7)
University : _____ (8)
Crowd-sourced : _____ (11)
Data is created / collected by data collaborative : _____ (12)
Other : _____ (10)
Total : _____

Q12 Are there any other data collaboratives, data pools, or general collaborations that inspired the data collaborative? If so, please specify whom.

○ Yes (4) _____

○ No (5)

End of Block: Organizational Governance

Start of Block: Measures of Success

Q13 Please give a list of the goals of your organization. Please just list items, no long explanation is required.

_____

_____

_____

_____

_____

Q14 What challenges do you face in your work? Please check all that apply.

	Significant challenge (1)	Minor Challenge (0)
Funding (4)	○	○
Sufficient technical resources (5)	○	○
Bureaucratic limitations (6)	○	○
Guidance/Knowing how (9)	○	○
Time/Deadlines (10)	○	○
Lack of stakeholder support (13)	○	○

Finding volunteers (12)	○	○
Lack of data collaborator support (14)	○	○
Other (11)	○	○

Q15 Overall, and looking over the time your data collaborative has been in existence, how successful have you been on the following scale: From 0 (failure) to 7 (we have met all of our objectives), and 10 (we are successful also in ways that we did not anticipate or plan).

	0	1	2	3	4	5	6	7	8	9	10
How successful are you? ()											

Q16 Are there any recommendations you would like to make to other data collaboratives, including those that are just starting, or those already in existence for some time?

_____

_____

_____

_____

_____

End of Block: Measures of Success

Start of Block: Conclusion

Q17 Which title best identifies your role within the collaborative? Please select the best answer.

- ○ Board Member (4)
- ○ Director (5)
- ○ Program Manager / Project Manager (6)
- ○ Software Engineer (7)
- ○ Data Analyst (8)
- ○ Participant / Member (9)
- ○ Elected Official (10)
- ○ Liaison / Ambassador (11)
- ○ Citizen (13)
- ○ Other (12) _____

Q18 Can we contact you for further follow-up?

- ○ Yes (1)
- ○ No (2)

Skip To: End of Survey If Can we contact you for further follow-up? = No

Q19 Can we obtain your governance charter? (e.g., Charter, Municipal Code, Organizational By-Laws)

- ○ Yes (1)
- ○ Maybe (2)
- ○ No (3)

Q22 Would you like us to update you with results of this research?

- ○ Yes (4)
- ○ No (5)

Q20 Name

_____

Q21 Email Address

_____

End of Block: Conclusion

Appendix C: Survey Reminder Emails

EVALUATING THE GOVERNANCE AND PERFORMANCE OF DATA COLLABORATIVES

RECRUITMENT SCRIPTS

1. INITIAL CONTACT 2-3 DAYS BEFORE SURVEY LINK EMAILED

Dear [insert name],

We are writing to ask for your participation in an important research study conducted at the Ohio State University to better understand the management and governance of data collaboratives. Despite the rise in these new ways to collect, use and distribute information, we have little systematic research on how data collaboratives are actually governed and managed.

Over the last few months, our research team spent many hours identifying data collaboratives, and this has led us to you. In the next few days, you will receive an email from us with a link to a survey on data collaboratives; it will take about nine minutes to complete.

The survey is designed to be answered by the person who "best understands how the data collaborative is governed and managed." If you do not think that you are the best person to answer this survey, please reply to this email with contact information for the appropriate city personnel.

If you are concerned that this is a legitimate survey, you can visit my web page and contact me by email or calling my cell phone at 614.795.6002.

Sincerely,

David Landsbergen, PhD, Associate Professor
614-292-9577 Office
landsbergen.1@osu.edu

John Glenn College of Public Affairs
The Ohio State University
1810 College Road
Columbus, OH 43210

2. SECOND CONTACT, FIRST LINK TO SURVEY

Dear [insert name],

A few days ago, we emailed you to ask for your participation in an important research study conducted by The Ohio State University to better understand the management and governance of data collaboratives. Data collaboratives are a new way to organize efforts to collect and use information.

We need your help in collecting information about how data collaboratives are governed and managed with the goal of sharing those results with others. While great strides have been made in data collaboratives, there is little empirical research on how data collaboratives initiatives are actually managed and governed.

The survey will take approximately nine minutes to complete and can be accessed here:
Take the Survey
Or copy and paste the URL below into your internet browser:
[Link]
If you are concerned that this is a legitimate survey, you can visit my web page and contact me by email or calling my cell phone at 614.795.6002.

To ensure a representative sample, we need the requested information by [insert date apx 2 weeks from date sent].

We are looking for the person who "best understands the governance and management of data collaboratives." If this does not apply, please reply to this email with the correct contact information

We very much look forward to hearing from you.

Sincerely,

David Landsbergen, PhD, Associate Professor
614-292-9577 Office
landsbergen.1@osu.edu

John Glenn College of Public Affairs
The Ohio State University
1810 College Road
Columbus, OH 43210
Follow the link to opt-out of future emails:
Click here to unsubscribe

3. FIRST REMINDER, 1 WEEK AFTER SURVEY SENT

Dear {insert name}:

Last week we reached out to you and extended an invitation to participate in a brief survey designed to better understand the implementation and sustainability of data collaboratives. We continue to ask for your support in completing the survey because there is so little research on how data collaboratives are structured, funded, and governed.

Without your insights, the survey might not be representative or capture the issues that are most important to you. We are also committed to using the results from the study to inform policymakers and administrators with recommendations to address present challenges in data collaboratives.

Follow this link to the Survey:
Take the Survey
Or copy and paste the URL below into your internet browser:
[Link]
If you are concerned that this is a legitimate survey, you can visit my web page and contact me by email or calling my cell phone at 614.795.6002.

To ensure a representative sample, we need the requested information by [insert date apx 1 weeks from date sent].

We are looking for the person who "best understands the governance and management of data collaboratives." If you do not think that this describes your duties, please reply to this email with contact information for the appropriate city personnel.
Thank you for sharing your time and insights.
Sincerely,
David Landsbergen, PhD, Associate Professor
614-292-9577 Office
landsbergen.1@osu.edu

John Glenn College of Public Affairs
The Ohio State University
1810 College Road
Columbus, OH 43210

Follow the link to opt-out of future emails:
Click here to unsubscribe

4. SECOND REMINDER, 2 WEEKS AFTER SURVEY SENT

Dear {insert name}:

Earlier this month, we reached out to you and extended an invitation to participate in a short survey to better understand the implementation and sustainability of data collaboratives. We know your time is valuable, but without your insights, the survey remains incomplete.

The survey takes approximately nine minutes to complete. You can access the survey by using the following link:
Take the Survey
Or copy and paste the URL below into your internet browser:
[Link]
If you are concerned that this is a legitimate survey, you can visit my web page and contact me by email or calling my cell phone at 614.795.6002.

To ensure a representative sample, we need the requested information by [insert date apx 2 two days before survey closes].

We are looking for the person who "best understands the governance and management of data collaboratives." If you do not think that this describes your duties, please reply to this email with contact information for the appropriate city personnel.

Thank you for sharing your time and insights.
Sincerely,
David Landsbergen, PhD, Associate Professor
614-292-9577 Office
landsbergen.1@osu.edu

John Glenn College of Public Affairs
The Ohio State University
1810 College Road
Columbus, OH 43210

Follow the link to opt-out of future emails:
Click here to unsubscribe

5. FINAL REMINDER, 3 WEEKS AFTER SURVEY SENT

Dear {insert name}:

Last week we sent you a reminder to please participate in a short survey designed to capture your experiences designing and implementing data collaboratives. The window to complete the survey is closing on {insert date} and we want to ensure your voice is heard. The information you share will advise policymakers and administrators across the U.S. in making decisions about the sustainability and effectiveness of data collaboratives.

The survey takes approximately nine minutes to complete. You can access the survey using the following link:

Take the Survey
Or copy and paste the URL below into your internet browser:
[Link]

If you do not think that you are the best person to answer this survey, please reply to this email with contact information for the appropriate city personnel.
Thank you for sharing your time and insights.
Sincerely,
David Landsbergen, PhD, Associate Professor
614-292-9577 Office
landsbergen.1@osu.edu

John Glenn College of Public Affairs
The Ohio State University
1810 College Road
Columbus, OH 43210
Follow the link to opt-out of future emails:
Click here to unsubscribe

Appendix D: Follow Up Interview Questions

Results and Recommendations Provided Through the Survey

Here are some of our results and working conclusions. Let's talk about the results that are most salient to you.

1. The mean number of employees working in a data collaborative were n= 9

2. The mean number of volunteers working in a data collaborative were n=2 (only 30% of the collaboratives have volunteers

3. Receiving federal grants was associated with hiring more employees.

4. The funding sources are:

Funding	Average
Local government revenue	7%
State grant	4%
Federal grant	17%
Crowd Funded	2%
Public Private partnership	3%
Privately funded	3%
University	7%
Subscription / membership fees	9%
Foundation support	30%
Fee for Service	11%
Other	8%

5. Most of the data collaboratives responding to us were nonprofit with a small majority of public sector collaboratives.

6. Data Collaboratives are involved in the full spectrum of data collaborative activities. There was not much variance across the sample on this. In other words, most data collaboratives, large and small, public or nonprofit, performed most of these activities.

- Collect data
- Analyze data
- Disseminate data
- Disseminate analysis
- Work with Clients on Projects
- Curate data
- Work with Citizens on Projects
- Clean data
- Create ways to combine data

7. Most of the data collaboratives identified their main activities as:

Main activity	Percentage
Data pooling	17.39%
Prizes & Challenges	0.00%
Research Partnerships	8.70%
Intelligence Products	30.43%
API	0.00%
Trusted Intermediary	30.43%
Other	8.70%

8. The most significant MAJOR barriers to data collaboratives are (in order of importance):

Funding

Bureaucratic limitations

Time and Deadlines

9. The most mentioned MINOR barriers are (in order of importance):

Sufficient technical resources (50% of sample)

Guidance and know how to do data sharing

Time and Deadlines

The less important barriers were:

Lack of stakeholder support

Finding volunteers

Lack of data collaborator support

10. The most significant experiences / recommendations provided by your peers were:

· With more data now online, there is now a move towards more analysis of that data.

· You must work closely to understand the needs of your stakeholders.

· The significant work in running a data collaborative is less about technology and data and more about focusing and developing soft skills in building trust and developing relationships.

· Privacy must remain a priority.

· Maintain conversations with other data collaboratives

Appendix E: Survey Analysis Codebook

Data Collaboratives Environmental Scan Survey
Data Cleaning Codebook

Survey Metadata

Columns A-D contain irrelevant information such as start date, end date, response type, and IP address.

- These columns were hidden.

Column E refers to the percentage of progress completed.

- All rows with <100% were deleted.

Column F refers to duration in seconds.

Columns G-I contain irrelevant information such as finished (inferred from progress), recorded date, and response ID.

- These columns were hidden.

Column J-L contain Recipient Names and email as defined in our distribution contact sheet.

Columns M-R contain irrelevant information such as ExternalReference, Coordinates, Distribution Channel, User Language, and Informed Consent (implied by survey completion being 100%).

- These columns were hidden.

Survey Questions

Column S - Q3: Approximately, how many full-time equivalent employees do you have? -->fulltime

Column T – Q4: Approximately, how many full-time equivalent volunteers do you have? -->volunteers

Column U – Q5: In what year was your data collaborative formed? (YYYY) -->year

Column V – Q6: What best describes the main activities of the collaborative? Please pick the best answer. - Selected Choice -->mainactivity

- These options were converted into numerical values:
    - Data pooling -> 1 (Organizations agree to create a unified presentation of datasets as a collection accessible by multiple parties)

    - Prizes & Challenges -> 2 (Organizations make data available to participants who compete to develop apps; answer problem statements; test hypotheses and premises; or pioneer innovative uses of data for the public interest and to provide business value.)
    - Research Partnerships -> 3 (Organizations engage directly with public-sector partners and share certain proprietary data assets to generate new knowledge with public value.)
    - Intelligence Products -> 4 (Organizations internally develop data-driven analyses, tools, and other resources, and release those insights to the broader public.)
    - API - > 5 (Organizations provide open access to certain data assets, enabling independent uses of the data by external parties.)
    - Trusted Intermediary -> 6 (Third-party actors support collaboration between private-sector data providers and data users from the public sector, civil society, or academia.)
    - Other -> 7
- Column W contains “other” responses

Others	Recode as
Combination of data pooling, intelligence products, and APIs	
due to a proliferation of online data tools in greater boston over recent years we've shifted to function more like a data-driven think tank, producing articles, reports and white papers, than as a traditional data intermediary	Intelligence products
Citizen science project where volunteers contribute research quality data and the organization makes that data as accessible as possible to domain specific research.	Trusted intermediary
quality improvement technical assistance- data collection, data analysis, data reporting used in learning collaboratives	

Column X – Q7: What kinds of activities do you engage in? (Please check all that apply.)

- These options were converted to numerical values --> Activities
    1. Collect data -->actcollect
    2. Analyze data -->actanalyze
    3. Disseminate data -->actdisdata
    4. Disseminate analysis -->actdisanalysis
    5. Work with Clients on Projects -->actclientproj
    6. Curate data -->actcurate
    7. Work with Citizens on Projects -->actcitizenproj
    8. Clean data -->actclean
    9. Create ways to combine data -->actcreate
    10. Other -->actother
- Column Y contains “other” responses

Others	Recode as...
build tools for utilizing and understanding data	
Integrate with storytelling and policy analysis	
Technical assistance and capacity building	
data literacy capacity building for nonprofits	
Technical assistance to help organizations use their internal data	

Column Z – Q8: Which sector best describes your group? - Selected Choice

- These options were converted to numerical values --> sector
    1. Public
    2. Private
    3. Public Private Partnership (PPP)
    4. Non-profit
    5. Other
- Column AA contains “other” responses

Others	Recode as
Higher Education	
State Agency	Public
University research center	
Academic/Public Partnership	
Low-profit Limited Liability Company	Private

Column AB begins Q9: “What are your sources of funding for the data collaborative? Please enter the approximate percentage to the right of each revenue source.”

- The percentages correspond to the following columns:
    - Local government revenue – AB --> fundreve
    - State grant – AC -->fundstate
    - Federal grant – AD --> fundfederal
    - Crowd funded – AE --> fundcrowd
    - Public/private partnership – AF --> fundppp
    - Privately funded – AG --> fundprivate

    - College/University - AH --> funduniv
    - Subscriptions or Membership Fees – AI -->fundfee
    - Foundation Support – AJ -->fundfoundation
    - Other – AK --> fundother
    - Fee for service- (Created) – fundfeexser
- Column AL contains “other” responses

Others	Recode as
Fee for service, development and program revenue	Fee for service
fee for service	Fee for service
State general funds	
project-based funding public and private sources	
Various grants	
Donations from data users	Fee for service
Earned revenue from consulting	Fee for service
Grants/contracts	
Project Support	

Column AM – Q10: Who are the stakeholders in your work? (Please check all that apply.)

- These options were converted to numerical values: -->stakeholder
    1. Public Entity/Governmental Body --> shgov
    2. Citizens -->shcitizen
    3. Private Entity/Company -->shprivate
    4. Non-Profits -->shngo
    5. University -->shuniversity
    6. Research Group -->shresearchg
    7. Public -->shpublic
    8. Customers -->shcustomer
    9. Other -->shohter
- Column AN contains “other” responses

Others	Recode as
Community and decision makers from the public, private, and non-profit sectors.	Public – Non profits - Private
Legislators, practitioners, other state agencies, workforce boards	Public

Column AO begins Q11: “Whom do you collect data from? Please enter the approximate percentage to the right of each data source.”

- The percentages correspond to the following columns:
    - Public Entity/Governmental Body – AO --> datapublic
    - Citizens – AP --> datacitizen
    - Private Entity/Company - AQ --> dataprivate
    - Non-Profits – AR --> datango
    - University – AS --> datauniv
    - Crowd-sourced – AT --> datacrowd
    - Data is created / collected by data collaborative – AU --> datacollab
    - Other – AV --> dataother
- Column AW contains “other” responses

Column AX – Q12: Are there any other data collaboratives, data pools, or general collaborations that inspired the data collaborative? If so, please specify whom. - Selected Choice -->inspiration

- Column AY contains “other” responses

Column AZ – Q13: Please give a list of the goals of your organization. Please just list items, no long explanation is required. --> goals

Column BA begins Q14: “What challenges do you face in your work? Please check all that apply.”

- These columns correspond to the following challenges:
    - Funding – BA --> challfund
    - Sufficient technical resources – BB -->challtech
    - Bureaucratic limitations – BC --> challbureau
    - Guidance/Knowing how – BD -->challknowhow
    - Time/Deadlines - BE --> challtime
    - Lack of stakeholder support – BF --> challstakesup
    - Finding volunteers – BG -->challvolunteer
    - Lack of data collaborator support – BH --> challcollsup
    - Other – BI --> challother
- The following choices were converted into numerical values:
    - Minor Challenge – 1

    - Significant Challenge – 2
- Column BJ contains “other” responses

Column BK – Q15: Overall, and looking over the time your data collaborative has been in existence, how successful have you been on the following scale: From 0 (failure) to 7 (we have met all of our objectives), and 10 (we are successful also in ways that we did not anticipate or plan). - How successful are you? --> success

Miscellaneous Data

Column BL and beyond contains miscellaneous data that can’t be quantitatively evaluated. The second sheet, Misc. Data, contains:

- Q16: Are there any recommendations you would like to make to other data collaboratives, including those that are just starting, or those already in existence for some time?
- Q17: Which title best identifies your role within the collaborative? Please select the best answer.
    - These responses were converted to numerical values:
        1. Board Member
        2. Director
        3. Program Manager / Project Manager
        4. Software Engineer
        5. Data Analyst
        6. Participant / Member
        7. Elected Official
        8. Liaison / Ambassador
        9. Citizen
        10. Other
    - Column C contains “other” responses
- Q18: Can we contact you for further follow-up?
- Q19: Can we obtain your governance charter? (e.g., Charter, Municipal Code, Organizational By-Laws)
- Q22: Would you like us to update you with the results of this research?
- Q20: Name
- Q21: Email Address